DOCUMENT RESUME

ED 172 208 CS 204 859

AUTHOR Loux, Ann Kimble
TITLE Third Party Evaluation in a Cross Disciplinary Writing Program.
PUB DATE Apr 79
NOTE 14p.; Paper presented at the Annual Meeting of the Conference on College Composition and Communication (30th, Minneapolis, Minnesota, April 5-7, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Composition (Literary); *Expository Writing; *Grading; Higher Education; Inservice Teacher Education; *Interdisciplinary Approach; Minimum Competency Testing; *Student Evaluation

ABSTRACT

A group evaluation program of student papers was established within a cross-disciplinary undergraduate writing program. The program was designed to train students to write for a general audience rather than for one teacher, and was initiated within the context of an institution that required students to display objective writing competency in any one of several disciplines as a graduation requirement. The program sought to achieve these specific goals: agreement on competency standards among writing evaluators, separation of the evaluation of content from the evaluation of writing competency, and the extension of the audience for student papers. Institution-wide valid and reliable standards for competency were set, and faculty members were trained how and how not to teach writing. The following evaluation priorities emerged: strong emphasis should be placed on logic, coherence, development of ideas, and use of quotations; grammatical considerations should weigh less heavily than logic and coherence; structural, logical, and grammatical considerations are more significant than semantic or syntactical concerns; and no solution was found to the priority to be placed on poor writing. (DF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Ann Kimble Loux

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED172208

THIRD PARTY EVALUATION IN A CROSS DISCIPLINARY WRITING PROGRAM

Ann Kimble Loux

Most of us have strong intuitions or concrete verification that at least one of the claims students make about writing teachers is valid: evaluation, specifically grading, varies a great deal from teacher to teacher. "She doesn't like my style," always hits a tender nerve as writing styles do indeed provoke varied positive and negative aesthetic responses.

Certainly those of us who have worked in group "grading" sessions know how idiosyncratic individual evaluation is: repeatedly the same paper receives every grade from best to worst comparatively. Writing research has shown that a student's grade depends more on who graded the paper than upon what was written. In one experiment, there was strong evidence that when a teacher graded a group of papers, put them aside, and "then reread them some months later without reference to the grade given originally, there were significant discrepancies in the two sets of grades."[1] Such variety may give life spice and reinforce the existentialist's confidence that little is predictable, but it is also patently unfair to students, particularly when they are required to develop "objective" competency. If competency is or becomes a college requirement, clearly there should be no dispute over what competency is. The first goal, therefore, in third party evaluation of writing is agreement among evaluators.

One of the main reasons for my own conviction that writing should be taught within a discipline is that at the stage of assigning final grades (as well as during the course of the semester) criticism and evaluation should be separated. If the course is specifically a writing course, then the grade must be based solely on the student's writing, a difficult assessment to make because of the

CS204 859

personal nature of the written product. If, however, writing is taught in a content course, the objective, comparative hard and cold measure, the grade, can be given for the more easily determined content mastery. Writing competency can then be assessed separately on a simple pass/fail basis. The second goal in third party evaluation is, then, the separation of content mastery evaluation from writing competency evaluation.

A third goal of group evaluation in a multidisciplinary writing program that hopes to train students to write for a general audience is interdisciplinary selection and representation among the evaluators.

The test of a group evaluation approach lies, obviously, in the possibility of its practical application. And, while one swallow does not make a summer, one specific experiment offers some opportunity to analyze the problems involved in attempting to achieve these goals: 1) agreement on "competency" among writing evaluators, 2) separation of the evaluation of content from the evaluation of writing competency, and 3) extension of the audience for student papers.

St. Mary's College, a small Catholic women's liberal arts college in South Bend, Indiana, has been struggling through the development stages of a multidisciplinary writing competency program for four years. With competency a graduation requirement, a number of core courses had been designated "W Seminars" to enable students to gain a certification of competency, the "W." Although several teachers made themselves notable exceptions by maintaining strict standards, the general tendency was to award the "W" virtually automatically at the end of the semester. The implication that all, or most students developed into competent writers after one semester of instruction was simply not valid. Many of us dared not let the eye see the hand that declared some students to be "competent" writers. And of course those who had long upheld strict standards made those who "let everyone through" squirm. The first goal of

developing valid, reliable standards of evaluation was obviously not being met, and became a matter of serious concern.

The fact that writing competency was a separate requirement for graduation, not expressly connected with any single course, and that certification could be granted in any of several core courses of the curriculum—philosophy, psycology, sociology, government, history, humanistic studies, music—did enable the evaluators to separate content and writing evaluation from the beginning. The grade in the course was automatically separated from writing competency evaluation/certification, a stamp on the student's permanent record (only positive stamps appear on these records.) Goal number two was accomplished at the level of definition of the college writing proficiency requirement.

The third goal of developing a wide audience for multidisciplinary writing classes (an excellent indirect means of breaking out of the teacher-as-only-audience bind) and the first goal of developing valid, reliable standards for granting writing proficiency certification could be simultaneously achieved by committee rather than individual determination of the individual student's demonstration of proficiency in writing. On an experimental basis, an interdisciplinary committee of writing teachers embarked on the difficult and time-consuming enterprise of group evaluation.

Here, we must digress to look carefully at two concepts basic to any discussion of the evaluation of writing: reliability and validity. If evaluation is to be reliable, all evaluators must agree upon what is good writing and what is not and must apply standards universally. Education Testing Service's Paul Diederich's definition of reliability is "the amount of agreement between two sets of independent measures of the same characteristic, taken at about the same time."[2] If evaluation is reliable each evaluator will give essentially the same grade to each paper evaluated.

Reliability has been the central concern for evaluators in the past, but recent thought and research put more emphasis on validity. To be valid, the evaluation of writing must actually test the quality of writing. In the past so called tests of writing have determined the student's grammatical knowledge, subject matter mastery, range and depth of experience, memory or thinking ability--in short, almost everything but his ability to write. Such tests reliably measure the subject's ability to use semicolons, to name and describe the women in the Odyssey or to count out the causes of the rise in the pigeon population in St. Louis. Reliability has been a measure of what could easily be measured rather than of what truly needed measureing.

While I believe that the goal of valid and reliable evaluation within a single institution is not only attainable, but necessary, I am openly willing, at the present time, to forgo hope of achieving highly reliable results from one institution to another. After the evaluators within a particular institution arrive at a set of standards they find to be widely acceptable, they will be conscious of the criteria they used in developing these standards. They might then be able to adapt these specific criteria to the particularities of different institutions, but it would be a wild shot in the dark to try and impose an externally established writing proficiency level upon any individual writing program. Students vary enormously from institution to institution and respect for the principle of neither under-estimating nor expecting the impossible of students demands that the writing teachers in each institution develop their own criteria for competency. The process of determining the goals of a writing program, of defining competency, is itself helpful on two counts: 1) in developing institution-wide valid and reliable standards for competency, 2) in training teachers how and how not to teach writing.

The St. Mary's College group evaluation experiment was fortunate in that

the committee had only to determine competency, basically a pass/fail judgment; research suggests that when the task is thus simplified to basically a single judgment, agreement increases among evaluators.[3] The idea of group evaluation was taken up, in the beginning, by teachers who had volunteered to teach writing but who were self conscious about their lack of specialized training and experience. The teaching of writing is, of course, an area in which few outside or, for that matter, inside English departments have formal training, and also an area which English department members may jealously guard as a private preserve. Other faculty members often have little confidence in their own ability to teach writing; English faculty members often have still less confidence in the ability of anyone outside the disciplinary fraternity. In this situation, it was our experience that participation in the process of group evaluation developed the instructors' ability to teach writing effectively and to trust their own ability.

About half of the Writing Proficiency staff, eleven teachers, volunteered for a group evaluation experiment last semester (somewhat to my surprise, six of them were English department members). Throughout the semester the group worked toward a common understanding of objective standards, primarily by discussing a series of "borderline" papers (as judged by the teacher who had volunteered a set at any given session. A teacher would select five medium to poor student essays on an assigned topic (perhaps an analysis of imagery in a Shakespearean sonnet from an Introduction to Literature course, or a review of strategies of musical selection at a consert, from on Introduction to Music section), have the papers duplicated, and give copies to all other member of the evaluation group. The other evaluators would, independently of each other, rank the papers from best to worst, and predict W certifiability status (now? later? ever?) for each student writer. Then the group would meet, share rank-

ings, and discuss the how's and why's of the members' judgments. As expected, in the first practice session two papers received every comparative ranking from best of the lot to worst.

Early in the evaluation process, the committee began to formulate a simple, but as it turned out, far-reaching set of priorities:

1. Strong emphasis should be placed on logic, coherence, the full development of ideas, and the proper use of quotations as supporting evidence.
2. Grammatical considerations should not weigh as heavily as logic and coherence. A paper should not be rewarded for being merely grammatically correct.
3. Diction problems and/or awkward sentence constructions pose severe and frequent evaluation headaches because of their tenacious quality--can such errors be eradicated in a lifetime, much less in one semester? In general, structural, logical, and grammatical considerations are regarded as more important than semantic or syntactical ones.
4. The committee was unable to agree on where, in terms of priorities, to place undeveloped, immature, vacuous writing. In the end such papers gave the committee its greatest difficulty, and rightly so, as such writing routinely presents the greatest problems for most teachers at the college.

The basic consideration underlying the above scale of emphasis is the distinction between impossible and possible worlds. The committee ranked these elements of prose not merely according to importance in composition but, more importantly, according to the degree in which student papers are susceptible to improvement. Quite honestly, while we can hope for relatively rapid progress in organization and correctness of punctuation, experience suggests that students need to immerse themselves in the written dialect (and in life!) for years before they are capable of expressing complex ideas in complex grammatical forms.

After the committee's second and third sessions, it became clear that group evaluation was more demanding, "tougher," then individual evaluation; several of the evaluators expressed concern that the organized procedure of committee evaluation would considerably increase the number of those denied Writing Proficiency certification. One even suggested that when teachers get together

and "show off their high standards to one another," students invariably suffer.

At semester's end, each of the 11 evaluation experimenters submitted his/her student's portfolios for group evaluation. Two (in some difficult cases, four) readers other than the class teacher evaluated each portfolio. Three selections from the student's work had been placed in the right-hand folder pocket: 1) one original paper from any point in the semester, plus its rewritten version if any; 2) the final formal paper of the semester; and 3) a thirty-minute unannounced in-class essay written in the final week of the semester. The rest of the semester's work was available in the left-hand pocket of the student's portfolio if readers needed further information. The evaluators read quickly, holistically, following the suggestions of Coffman, McColly, and Coward who had found that analytic scoring takes twice as long yet yields approximately the same validity as holistic evaluation.[4]

After completing a whole class's portfolios, the two readers consulated and then submitted their recommendations and reasons to the class teacher. The recommendations had to do entirely with "W" certification, whether that requirement for graduation had or had not been met by the student. The readers made no recommendations about course grades, which remained the responsibility of the teacher. In separating the course grade from the writing proficiency certification, the committee attempted insofar as possible, to make evaluation reliable to evaluate on the basis of intrinsic criteria, i.e., the quality of presentation of ideas, or "relative readability."

The evaluation group read a total of one hundred and fifty portfolios first semester and one hundred second semester and proferred recommendations on them. The course teachers (themselves also evaluators for other teachers' student portfolios), with only four exceptions, accepted these recommendations, although they were in no way bound to do so. Given a rise in W denials of

almost fifteen precent, the final consensus was, to say the least, remarkable.

* * *

The process of group evaluation itself was fascinating, but I would never have predicted just how effective a tool it became in dealing with the many major problems of an interdisciplinary program. Most importantly, the procedure helped the teachers in the program develop effective assignments and guidelines for responding positively and productively to student writing. As we read sample student papers and discussed them, we developed agreement on criteria for evaluating writing. Each teacher became more and more confident of her abilities to teach writing. When the whole group served as evaluators, the responsibility for individual decisions was shared by all.

Visualize a writing committee of four English teachers and one each philosopher, historian, political scientist, psychologist, musician, sociologist which has just been formed to integrate carefully planned writing assignments into an introductory course in each of the respective disciplines. In almost no time numerous problems intrude, subtly or overtly, into all planning sessions. The English teachers are fearful of losing their private territory, their own bread and butter or that of their graduate students. Cross disciplinary programs are a real threat to beleaguered English departments and it's unwise not to acknowledge this reality. In fact, if the English department is too beleaguered, such a program would probably work only with an advanced course beyond the required English writing courses. The English teachers may also be somewhat less than comfortable with the credentials and competency of the sociology or history teacher as a writing instructor. And the history or philosophy teacher may be less than absolutely confident about his abilities to teach writing, though the last confessor he might choose would

be an English teacher. The evaluation procedure was an enormous success on both fronts. English teachers had to see and admit that their successes with student writers were in no way greater than the government teacher's. Honestly, I would say that English literature, next to philosophy, is clearly the most difficult discipline in which to learn to write. In both these disciplines the students has first to deal with complex reading materials and then to write; the student who writes on the radification of the ERA in her state or conserving energy in her family home will probably not have as many complex mental tasks before her as the student who writes a character analysis of the young boy in "Araby." The first and most important effect of group evaluation was an increase in confidence among the non English faculty in their ability to teach writing effectively and a reciprocal, possibly begruding, often only tacid, admission in the English faculty that their colleagues in other disciplines were doing creditable work, and had truly become our allies.

Another touchy issue that was handled indirectly by the group evaluation process involved formulas for required writing instruction. English department members in the beginning of the program were more than willing to tell, in considerable detail, those they regarded as novice writing teachers exactly how to do it, even though they themselves often had little or no training except combat duty where they might even admit to repeated failures. After working with the evaluation committee, we found that the English department was listening as well as offering advice (as much as, occasionally). In the beginning the English staff had wanted to establish a formula which specified how much time should be spend formally teaching writing in every non-literature course as well as in English classes. The proposal which narrowly missed becoming a requirement in the writing program was that each teacher must spend one third of his class time formally teaching how-to's -- word choice, outlining, sentence construction and combining, audience awareness,

thesis statements. As we looked at the groups of five papers in each of our training sessions, one point emerged again and again: rarely would two of the five papers have similar problems--each paper was unique. The most effective way to handle both evaluation and how to instruction was, we decided, individually with extensive comments on individual papers, conferences with individual students. Many of us were at last able to find some relief from the guilt we had long felt over not ordering, or ordering and not using, how-to-write texts.

Another fringe benefit of working on the evaluation committee— we all found ourselves picking up any number of hints for developing assignments that worked. The colossal flops in every discipline were obvious. Our Greek literature teacher has given up asking students to compare Greek lyrics and Greek epics; that's like comparing sardines with whales. Instead of a term paper, a history teacher now assigns seven brief papers on such topics as "Which colony would your forefathers have chosen to immigrate to had they left Europe in the seventeenth or eighteenth centuries? Explain in detail their situation and the reasons for their choice." We even have a sworn oath from one Philosophy teacher not to repeat the assignment which stood our hair on end. (1) Describe the logical steps of Descartes' and Anselm's proofs for the existence of God. (2) evaluate the effectiveness of these proofs. (3) Then give your own proof for God's existence and evaluate the persuasiveness of your argument." (Two to four pages.)

The most important result of group evaluation has been that all the teachers on the committee, from six different disciplines, have come to see themselves as sharing the responsibility for and anguish over the failures we all experience in our attempts to teach students to write. Not one of us need fear the question, "Why on earth did you give Eudora Welty her "W?"

Fear of being asked this question, a question which implies complicity in the assumption that writing miracles occur overnight, has had far too great an impact on the teaching of writing.

Now just a couple of comments on the effects of group evaluation of the teaching of writing generally. Group evaluation develops wide audiences for student writers and helps teachers to separate content mastery evaluation from writing evaluation.

If one aim in teaching writing is to help students internalize and converse with a general audience, then the concrete audiences who read their work would be as general as possible, in terms of both immediate criticism and long-range evaluations. Students badly need to move to issues more fundamental and objective than "He just doesn't like me," or "She doesn't appreciate my style," or "I get A's from her. Why does he give me C's?" The wider the audience reading their work, the less opportunity students have to dismiss criticism or grading as personal, idiosyncratic. If evaluation of their writing is fair and consistent (valid) student writers will come to respect the standards of the program at each institution.

One of the main reasons for my own conviction that writing should be taught within a discipline and not in a composition course, except remedially, is that at the stage of assigning final grades, content mastery and writing ability evaluation can actually be separated. If the course is specifically a writing course, then the grade must be based solely on the student's writing, a notoriously complex and touchy assessment to make largely because of the personal nature of the written product. If, however, writing is taught in a content course, the objective, comparative hard and cold measure, the grade, can be given for the more easily determined content mastery. In a cross disciplinary program writing competency can be assessed separately from

content mastery on a simple yes/no basis. Of course even the pass/fail evaluation is threatening to students, but at least everything isn't riding on a single talent--writing ability--an all or nothing proposition for skilled and unskilled writers, respectively.

I am somewhat uncomfortable with the argument given by Diederich that with group evaluation the class teacher and student form a useful league against "them" the evaluators. This league may in fact force the class teacher to assume a false position. She may find herself laying blame on "them" for opinions she helped "them" evolve and with which she thoroughly agrees. She herself is also "them" when she reads portfolios from other classes. Nevertheless, when I meet with my students to help them select their most creditable work for portfolio evaluation, I find that we are somehow "together"--invariably I hope for them and they know this well. While the other readers are going through my students' portfolios I can scarcely keep away from them, being curious every minute to know how my students are faring. There's a league, if not us against them, at least of all of us together hoping for each other's success, and occasionally sharing disappointments. I've never experienced a situation in which it is so clear to me that I care about what happens to my students.

Group evaluation has the same unsettling effects for both students and teachers as a good, hard look in a "nicely polished looking-glass." In the long run I believe group evaluation procedures can radically alter the course of writing teaching. Specifically, in my experience, the procedure works successfully with two of the greatest difficulties in a cross-disiplinary writing program: (1) teaching training, (2) agreement upon and application of valid and reliable standards for evaluating student writing.

Ann Kimble Loux
Department of English
St. Mary's College
Notre Dame, IN. 46556

FOOTNOTES

[1] William Coffman. "Essay Examinations." Educational Measurement, Washington, D. C.: American Council of Education, 1971, p. 291.

[2] Paul B. Diederich. Measuring Growth in English. Urbana, Illinois: National Council of Teachers of English, 1974, p. 36.

[3] Coffman, p. 259.

[4] Ibid.